Motif based Hessian matrix for *ab initio* geometry optimization of nanostructures

Zhengji Zhao,* Lin-Wang Wang, and Juan Meza

Computational Research Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720

(Dated: April 12, 2006)

Abstract

A simple method to estimate the atomic degree Hessian matrix of a nanosystem is presented. The estimated Hessian matrix, based on the motif decomposition of the nanosystem, can be used to accelerate *ab initio* atomic relaxations with speedups of 2 to 4 depending on the size of the system. In addition, the programing implementation for using this method in a standard *ab initio* package is trivial.

PACS numbers: 73.22.-f, 71.15.Mb, 79.60.Jv

*Electronic address: zzhao@lbl.gov

Atomic position relaxation is one of the most widely used features of a first principles *ab initio* calculation. With the maturization of the local density approximation (LDA) [1] and general gradient approximation (GGA) [2] of the density functional theory (DFT) [3], and with the advance of large scale supercomputers and related parallel codes, *ab initio* calculations can now tackle problems with hundreds or even thousands of atoms. Due to the recent interest in nanoscience, using *ab initio* methods to find the atomic structure of a given nanosystem has become a popular way of studying these problems [4]. For example, *ab initio* calculations have been used to determine atomic structures and catalytical properties of many atomic clusters on different substrate surfaces [5–7] and they have also been used to study the shapes of small nanowires and nanowire-molecule contacts [8, 9]. While there are various algorithmic developments aimed at speeding up the electronic structure calculation for a given atomic configuration, especially to make the computation scale linearly with the size N of the system [10], the methodology development for the atomic relaxation part is rare by comparison, especially for solid state systems.

There are a few commonly used algorithms for atomic relaxation in the state of the art DFT software packages, most noticeably: conjugate gradient (CG) method [11]; BFGS (Broyden-Fletcher-Goldfarb-Shanno) method [12]; RMM-DIIS (Residual MiniMization-the Direct Inversion in the Iterative Subspace) method [13]; and Damped Molecular Dynamics (DMD) method [14]. They are usually used without any preconditioning (no preconditioning for the CG and the DIIS, the initial Hessian for the BFGS set to the identity). While there are many studies in quantum chemistry for ways to speedup the geometry optimization of organic molecules, for example, by using internal coordinates [15–18] or by estimating the Hessian matrix using classical force field models [19–21], their applications to solid state systems are not trivial. Due to the large degree of redundancy, the application of internal coordinates (bond lengths, angles and torsion angles) to solid bulk and periodic systems has only been tested recently [22]. For the classical force field model Hessian matrix estimation, either the models are explicitly built for organic molecules [19–21], or the universal formula [23] might not work well for nanosystems as we will show below. Given this situation, the need for better atomic relaxation algorithms for solid state systems is apparent as we study larger systems (i.e., nanostructures with hundreds to thousands of atoms). As pointed out by Goedecker [24] *et al*: without an improvement in the atomic optimization method, the scaling of the computing time will not be $O(N)$ even if the electronic structure calculation is $O(N)$.

In this paper, we will present a simple method to accelerate the atomic position relaxation

of nanostructures for *ab initio* calculations. In the theoretical study of atomic configurations in nanostructures, one can use two different approaches. In the first approach, a molecular dynamics or Monte Carlo method is used to search for different bonding rearrangements and topologies. In the second approach, a given bonding topology is proposed, but the exact atomic positions and its total energy still need to be calculated using atomic relaxation. We will concentrate on the second approach. In this approach, although the final bonding topology is known, the system still spends many iterations near the minimum for relaxations. Thus a good estimate of the Hessian matrix of the atomic degree of freedom can potentially accelerate the relaxation process. Our method will be focused on estimating a good Hessian matrix. However, unlike the other works that estimate the Hessian matrix based on classical force field models [19–21], and depend on the existence of good parameters that are not always available, we build our Hessian matrix using a motif method. Although there are thousands of atoms, and thousands of degrees of freedom in a nanostructure, these atoms rarely represent thousands of bonding situations (hence are completely chaotic and disordered). Instead, the local bonding environments for those thousands of atoms are often similar (or repeating), and there are typically only a few local atomic bonding types. For example, in a colloidal crystalline, the atoms inside the quantum dot are all in a similar bonding type, while the atoms on the surface might have a few different passivation types. We will thus develop motifs, consisting of a few atoms, that are representative of these local bonding types. We then calculate the Hessian matrix elements of these motifs from *ab initio* calculations for the much smaller prototype systems, and then assemble these motif Hessian elements to approximate a Hessian matrix for a given nanostructure. The concept and procedure here is very similar to that of the motif based charge patching method [25, 26] for charge density calculations. That method has been successfully applied to many nanosystems, enabling the calculations of electronic structures of thousand atom systems. After this motif based Hessian matrix is calculated, it is used as a preconditioner in the CG method or as the initial Hessian matrix in the BFGS method (the two methods which will be focused on in this paper).

Let's assume there are N_m local atomic bonding types in a nanostructre. Note that, to be the same bonding type for two local bonding environments, we only require the nearest neighbor bonding atoms be the same, but not the same bond length and angles [25, 26]. After all, we only need to construct an approximate Hessian matrix to accelerate the convergence. Now, for each local bonding type, we will define a motif. This motif m consists of one center atom at $R(l_0)$ and its nearest neighbor bonding atoms at $\{R(l_b)\}$, where l_0 and l_b are the atom indices, and R denotes

the atomic position. We will then use the Hessian matrix elements of these motifs to approximate the Hessian matrix of the entire nanosystem. The Hessian matrix H is defined as

$$H(l,i;l',i') = \frac{\partial^2 E}{\partial R_i(l)\partial R_{i'}(l')} = \frac{\partial F_{i'}(l')}{\partial R_i(l)}, \tag{1}$$

here l and l' are the indices of atoms, i and $i' = 1,2,3$ denote the x, y, z Cartesian coordinates, $R_i(l)$ is the ith Cartesian coordinate of the atom l, E is the total energy of the system, and $F_i(l)$ is the ith component of the force on atom l.

To make a simple approximation, we will drop the Hessian matrix elements beyond the nearest neighbor atoms. That is, we set $H(l,i;l',i') = 0$ if atom l and atom l' are not connected by a bond. This is justifiable since the Hessian matrix elements between the next nearest neighboring atoms and beyond are typically an order of magnitude smaller than the ones between nearest neighboring atoms, as shown in Fig. 1. We have ignored possible long range electric field effects, but they are usually weak due to the screening. We now calculate $H(l,i;l',i')$ for the motif m. All we need here is the matrix elements between the center atom $R(l_0)$ and its nearest neighboring atoms $\{R(l_b)\}$: $H_m(l_0,i;l_b,i')$. This can be obtained easily via Eq. (1) by *ab initio* calculations of small prototype systems containing the motif, and by displacing the center atom with a small displacement $\Delta R_i(l_0)$ and calculating the forces at the neighboring atoms. For example, for a tetrahedral bulk system, like bulk Si and bulk GaAs, one motif consists of the center atom and the four nearest neighboring atoms, and an 8 atom cubic supercell can be used to calculate the motif Hessian elements.

After the required motifs are calculated, the Hessian matrix of a nanostructure can be constructed as follows:

$$H(l,i;l',i') = \begin{cases} H_m(l,i;l,i') & \text{if } l = l' = l_0 \text{ of motif } m; \\ \frac{1}{2}[H_m(l,i;l',i') + H_{m'}(l',i';l,i)] & \text{if } l,l' \text{ are neighboring atoms;} \\ 0 & \text{if } l,l' \text{ are beyond neighboring atoms.} \end{cases} \tag{2}$$

In the above equation, for the second case, m and m' are two neighboring motifs, and they have l and l' atoms at their centers respectively. Note, one can consider $H_m(l,i,l',i')$ as the i'th component of the force at the neighboring atom l' when the center atom l moves in the ith direction. Also, by construction, the Hessian matrix H is symmetric between the indices of $\{l,i\}$ and $\{l',i'\}$.

However, for systems with no external electric fields (e.g., isolated quantum dots and bulks), the total force on the system should be zero. This provides some additional constraints (sum rules)

on the Hessian matrix:

$$\sum_{l'} H(l,i;l',i') = 0 \text{ for all } l, i \text{ and } i'. \tag{3}$$

These sum rules are not automatically satisfied by a Hessian matrix constructed from Eq. (2). To better satisfy these sum rules while maintaining the symmetry of the matrix, we have done the following. If we define $\Delta H(l,i,i') = \sum_{l'} H(l,i;l',i')$, then we first modify the on-site ($l = l'$) H element as $H(l,i;l,i') = H(l,i;l,i') - \Delta H(l,i,i')$. This will guarantee that the sum rules 3 will be satisfied. Unfortunately except for $i = i'$, the symmetry of these on-site elements will now be violated, *i.e*, $H(l,i;l,i') \neq H(l,i';l,i)$. We thus symmetrize these on-site elements by setting $H(l,i;l,i') = [H(l,i;l,i') + H(l,i';l,i)]/2$. Although this will violate the sum rule again, the amplitude of this violation is much smaller. For example, in the quantum dot system to be discussed later, before the above operations, the sum rule violation as measured by $\sum_{l,i,i'} |\Delta H(l,i,i')| / \sum_{l,i,l',i'} |H(l,i;l',i')|$ is 2.0%, while after these operations, the sum rule violation is only 0.4%.

After the above treatment, $H(l,i;l',i')$ will be inverted and used as the preconditioner in the CG algorithm or as the initial inverse Hessian matrix in the BFGS algorithm. To invert it, we first diagonalize the matrix using $\{l,i\}$ as a compound index. Due to the force sum rules, there should be three zero eigenvalues corresponding to three translational modes. In our case, because the sum rules are not exactly satisfied, these three eigenvalues are not exactly zero. Nevertheless, they are about an order of magnitude smaller than the other eigenvalues (which are all positive). We constructed the inverse Hessian matrix H^I as $\sum_j \tilde{\varepsilon}_j V_j^T V_j$ where V_j is the eigenvector of H, and $\tilde{\varepsilon}_j$ equals the inverse of the original eigenvalues ε_j, except for the three smallest ones for which $\tilde{\varepsilon}_j = 0$. The resulting H^I is used in the CG or BFGS algorithms. Note that, in terms of the computational codes, the construction of H^I can be done completely outside the package before the actual *ab initio* atomic relaxations. The only requirement for any *ab initio* computational package is the ability to read the H^I and to use it in the CG or BFGS algorithm. For our tests, we have used the PEtot [27] LDA planewave pseudopotential package.

We first test the bulk Si and bulk GaAs. For bulk Si, there is only one motif needed, while for GaAs there are two motifs, one with Ga in the center, another with As in the center. All the motif Hessian elements are obtained from *ab initio* calculation of 8 atom cubic cells. We have used a 64 atom cubic supercell with random displacements within $\pm$ 0.1 a.u. from their ideal positions. Norm conserving pseudopotentials and 30 Ryd planewave cutoff are used. The *ab initio*

atomic relaxation results are shown in Fig. 2. The errors of the total energy $\Delta E = E - E_{\text{exact}}$ (measured from the exact solutions) are plotted as functions of atomic relaxation steps. For both CG and BFGS, within each step, after the search direction is obtained, one line minimization using quadratic approximation prediction is performed. From Fig. 2, we see that, without the use of our motif based Hessian, H_{motif}, CG is slightly better than BFGS. After H_{motif} is used, the preconditioned CG (denoted as H_{motif}-PCG) is similar to the H_{motif}-BFGS (the BFGS method with H^I_{motif} as the initial Hessian inverse), and they both are a factor of 2-3 faster than the unpreconditioned CG and BFGS. One possible reason for the plain BFGS being slower than CG could be our use of the identity matrix for the initial Hessian in the BFGS. Unfortunately, this is common practice in the implementation of the BFGS method. Note that, since our Hessian matrix describes the asymptotic behavior of the acoustic modes of the bulk systems correctly, one could expect larger speedups for larger systems [24].

The above tests for bulk systems are interesting, but are not very useful in practice since the exact solutions are known without doing *ab initio* calculations. We next test GaAs systems with N (nitrogen) impurities. Here, for comparison, we also test a universal classical force field model proposed by Fernandez-Serra *et al* [23]. In this model, a bond and bond angle energy terms from a force field expression are used for universal systems. This model was tested for gold clusters based on the embedded atom model. We have followed the exact procedure as described in Ref.23 to calculate the Hessian matrix from this model; the results for $Ga_{32}As_{31}N_1$ and $Ga_{32}As_{29}N_3$ systems are shown in Fig. 3. To start with, the atoms are in the ideal tetrahedral positions with additional random displacements within $\pm$ 0.1 a.u. As we can see, with the use of this model Hessian, PCG (denoted as H_{FF}-PCG), slightly improved the original convergence (by about 20%), while for BFGS (denoted as H_{FF}-BFGS), it is not stable as cautioned in their original paper [23]. Next, we used our motif method using four different motifs, one for a As center, one for a N center, and two for Ga centers with none or one N as a neighbor respectively. All the motif Hessian elements are calculated from 8 atom cubic cells. The convergence results of using our motif Hessian matrix, H_{motif}-PCG and H_{motif}-BFGS, are also shown in Fig. 3. We can see that the speedup for the convergence is similar to the bulk result, and is much better than the universal classical force field model.

All the above tests are for bulk systems, where there is no surface involved. We next test a Si quantum dot passivated by hydrogen atoms. On the surface, each Si dangling bond is passivated by one hydrogen atom. There are two surface bonding types: a Si atom with one hydrogen and a

Si atom with two hydrogen atoms. Adding the situation of bulk Si, there are three motifs with Si at the centers. The hydrogen centered motifs are trivial with their Hessian matrix elements derivable from the Si centered motifs using the symmetry and force sum rules. These surface motif Hessian matrix are calculated using an 18 atom (001) surface slab and a 32 atom (111) surface slab. We have used these three motifs to construct the Hessian matrix of a 375 atom Si quantum dot. The initial atomic positions of this quantum dot are estimated from a previous passivation model as discussed in Ref. 28. The computational results with and without using our motif Hessian for the CG method are shown in Fig. 4 (open and filled triangles, respectively). During this relaxation, the maximum atomic displacement is 1.46 a.u. (the average is 0.83 a.u.), measured by the difference between the initial atomic positions to their final converged positions. As we can see from Fig. 4, almost a factor of 4 speedup is achieved by using our new method. Note that this speedup for the quantum dot is slightly faster than that for the 64 atom bulk Si, even though a bulk system might be the more appropriate problem for this kind of approach. We attribute this to the fact that the quantum dot system is larger than the 64 atom bulk system, thus one would expect a larger speedup. In Fig. 4 we have also shown the results of another Si quantum dot with 339 atoms (open and filled spheres, respectively). In this quantum dot, the surface Si atoms with two dangling bonds form Si-Si dimers (between two parallel zigzag chains in the (110) direction). This is a common surface reconfiguration on a Si surface. As a result of this surface reconstruction, the quantum dot center Si-Si bond length is 0.3% larger than the previous un-dimerized quantum dot. As shown in the Fig. 4, the speedup for this quantum dot is similar to the previous quantum dot, showing the robustness of our method to different surface reconstructions. In the above test, we only compared our quantum dot results to the unpreconditioned CG method. We have also run the same quantum dot system with other algorithms, e.g., BFGS (using PEtot), and RMM-DIIS, DMD (using the VASP code [29]). We find that the speeds of BFGS, DMD, RMM-DIIS, are all very similar to the CG method (within 20%), they are all much slower than our preconditioned CG method.

Since *ab initio* atomic relaxations often take hours to converge even on large supercomputers, the time saved can be significant. For example, in above 375-atom Si quantum dot test, one iteration takes 140 minutes both for PCG and CG. We used 32 processors on Seaborg (NERSCs 6080-processor IBM RS/6000 SP has 380 16-CPU POWER3+ SMP nodes with 16G, 32G, or 64G memory; each processor has a peak performance of 1.5 GFlops), as we can see (Fig. 4), 17 iterations of CG and 5 iterations of PCG have reached the same convergence, that means 2380

minutes of CG vs. 700 minutes of PCG, the actual saving is 1680 minutes, counting the 32 processors, then the actual saving is 896 CPU hours on Seaborg.

In summary, we have presented a simple method to approximate the Hessian matrix of a nanosystem based on motifs. The use of such a Hessian matrix as a preconditioner in the CG and BFGS methods can speedup the convergence of atomic relaxations by a factor of 2 or 4, and the larger the system, the greater the speedup one can expect. Since *ab initio* atomic relaxations often take hours to converge even on large supercomputers, the time saved can be significant. As a by product, our method is always able to provide a good initial atomic positions through the *ab initio* atomic relaxations for the small prototype systems when calculating the Hessian elements for motifs. Like all other Hessian matrix methods, our method is applicable for a system for which the bonding topologies are known or selected. The implementation requirements for standard *ab initio* packages to use this method are minimal, since the Hessian matrix can be calculated outside the code.

This work was supported by the director of Office of Science, Office of Advanced Scientific Computing Research, Office of Basic Energy Science and U. S. Department of Energy, under Contract No. DE-AC02-05CH11231. This research used the resources of the National Energy Research Scientific Computing Center.

---

[1] W. Kohn and L.J. Sham, Phys. Rev. **140**, A1133 (1965).

[2] J.P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).

[3] P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964).

[4] A. Puzder, A.J. Williamson, J.C. Grossman, and G. Galli, J. Chem. Phys. **117**, 6721 (2002).

[5] S. Pick, V.S. Stepanyuk, A.N. Baranov, W. Hergert, and P. Bruno, Phys. Rev. B. **68**, 104410 (2003).

[6] M. Moseler, H. Hakkinen, and U. Landman,Phys. Rev. Lett. **87**, 53401 (2001).

[7] J. Li, X. Li, H.-J. Zhai, and L.-S. Wang, Science, **299**, 864 (2003).

[8] P. Sen, O. Gulseren, T. Yildirim, I.P. Batra, and S. Ciraci, Phys. Rev. B. **65**, 235433 (2002).

[9] J.W. Kang, and H.J. Hwang, Nanotechnology, **12** 295 (2001).

[10] S. Goedecker, Rev. Mod. Phys. **71**, 1085 (1999).

[11] J. Nocedal, and S.J. Wright, *Numerical Optimization* (Springer, New York, 1999).

[12] W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery, *Numerical Recipes* (Cambridge Uni-

versity Press, Cambridge, 1992).

[13] P. Pulay, Chem. Phys. Lett. **73**, 393 (1980).

[14] F. Tassone, F. Mauri, and R. Car, Phys. Rev. B **50**, 10561 (1994).

[15] P. Pulay and G. Fogarasi, J. Chem. Phys. **96**, 2856 (1992).

[16] P. Pulay and B. Paizs, Chem. Phys. Lett. **353**, 400 (2002).

[17] K. Nemeth and M. Challacombe, J. Chem. Phys. **121**, 2877 (2004).

[18] V. Bakken and T. Helgaker, J. Chem. Phys. **117**, 9160 (2002).

[19] H. B. Schlegel, Theoret. Chim. Acta **66**, 333 (1984).

[20] T.H. Fischer and J. Almlof, J. Phys. Chem. **96**, 9768 (1992).

[21] R. Lindh, A. Bernhardsson, G. Karlstrom, and P. Malmqvist, Chem. Phys. Lett. **241**, 423 (1995).

[22] K.N. Kudin, G.E. Scuseria, and H.B. Schlegel, J. Chem. Phys. **114**, 2919 (2001).

[23] M.V. Fernandez-Serra, E. Artacho, and J.M. Soler, Phys. Rev. B **67**, 100101(R) (2003).

[24] S. Goedecker, F. Lancon, and T. Deutsch, Phys. Rev. B **64**, 161102(R) (2001).

[25] L.W. Wang, Phys. Rev. Lett. **88**, 256402 (2002).

[26] L.W. Wang and J. Li, Phys. Rev. B **69**, 153302 (2004).

[27] http://hpcrd.lbl.gov/l̃inwang/PEtot/PEtot.html

[28] L.W. Wang and A. Zunger, in "*Nanocrystalline Semiconductor Materials*", Edited by P.V. Kamat and D. Meisel (Elsevier Science. Amsterdam, Netherland, 1996).

[29] http://cms.mpi.univie.ac.at/vasp/vasp/vasp.html

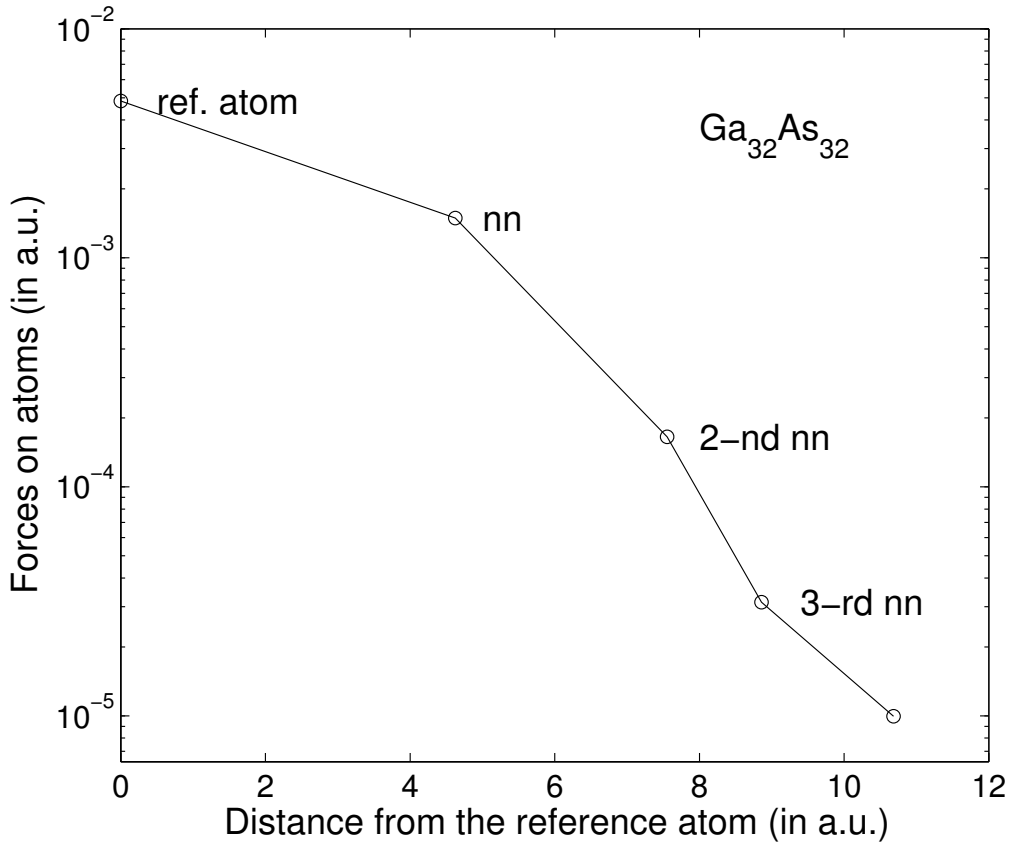

FIG. 1: The average forces on atoms when the reference atom (As) moves along x direction a small distance (around 0.04 a.u.) in $Ga_{32}As_{32}$ system, where nn represents the nearest neighbor.

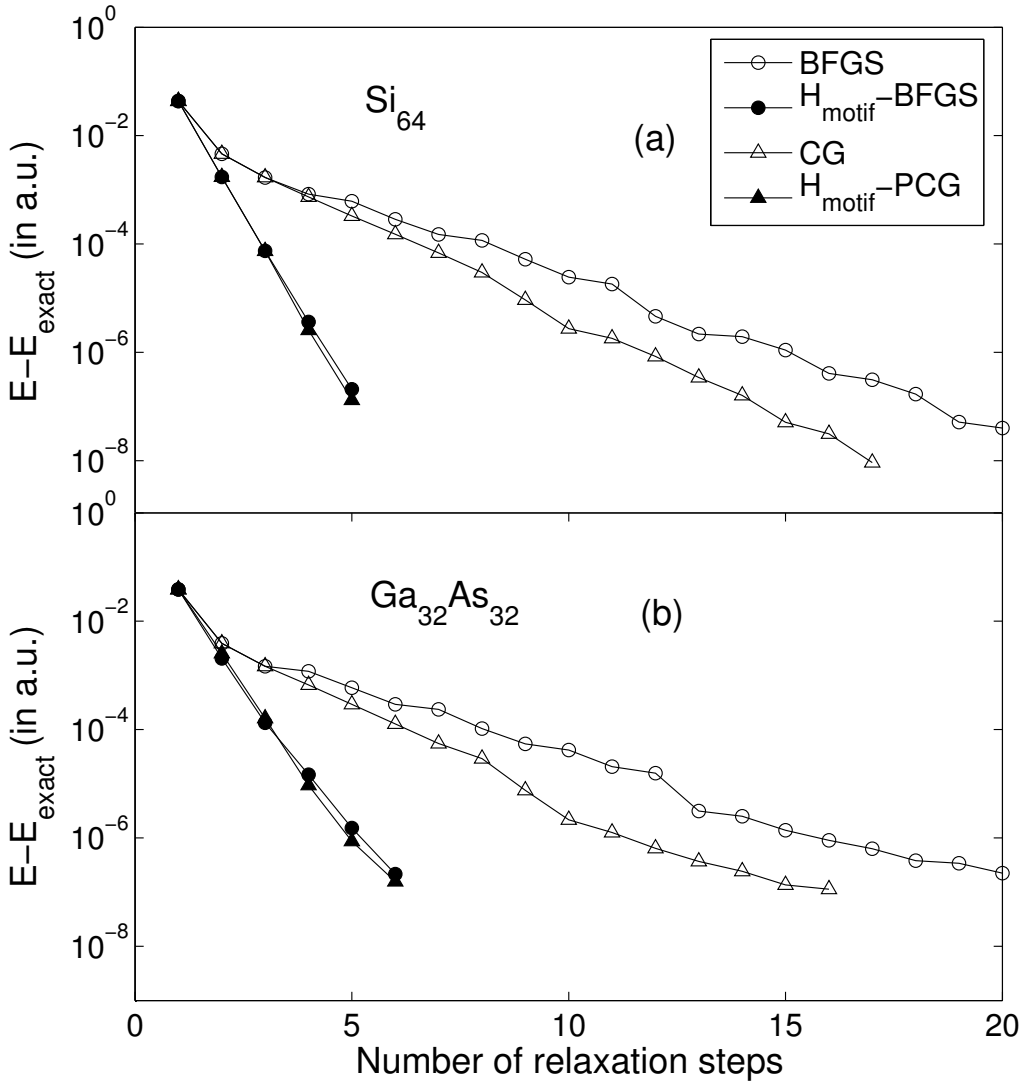

FIG. 2: The energy convergence in atomic relaxations for Si_{64} (a) and $Ga_{32}As_{32}$ (b) systems by applying BFGS (open sphere), CG(open triangle), BFGS starting with our motif Hessian $H_{\rm motif}$ (filled sphere), CG preconditioned by $H_{\rm motif}$ (filled triangle), respectively. The error of the total energy (vertical axis) are the differences from the exact energy.

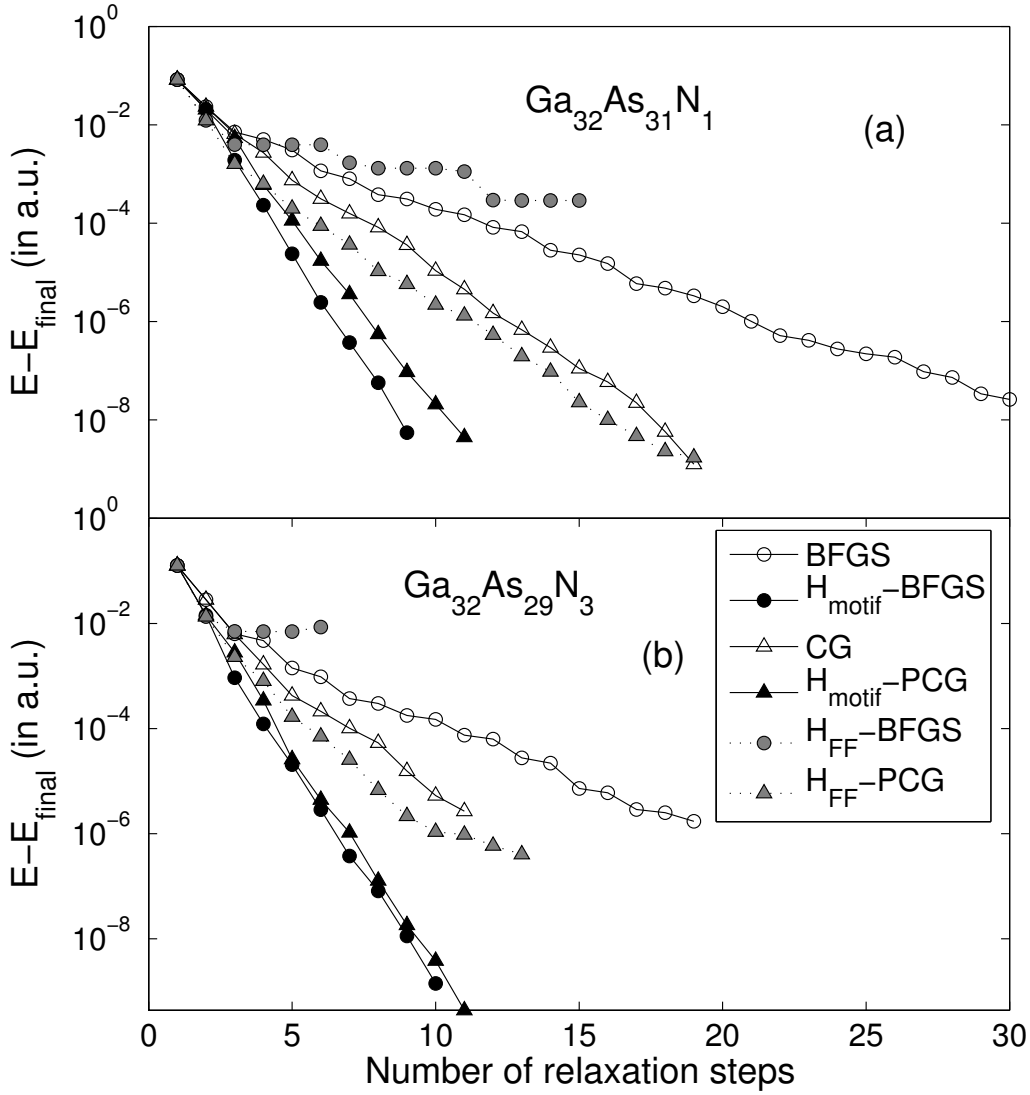

FIG. 3: The energy convergence in atomic relaxations for $Ga_{32}As_{31}N_1$ (a) and $Ga_{32}As_{29}N_3$ (b) systems by applying BFGS (open sphere), CG(open triangle), BFGS starting with our motif Hessian $H_{\rm motif}$ (filled sphere), CG preconditioned by $H_{\rm motif}$ (filled triangle), BFGS starting from the Hessian $H_{\rm FF}$ derived from the universal force field model from Ref. 23 (gray sphere), and CG preconditioned by $H_{\rm FF}$ (gray triangle) methods, respectively. The error of the total energy (vertical axis) are the differences from the final converged energy.

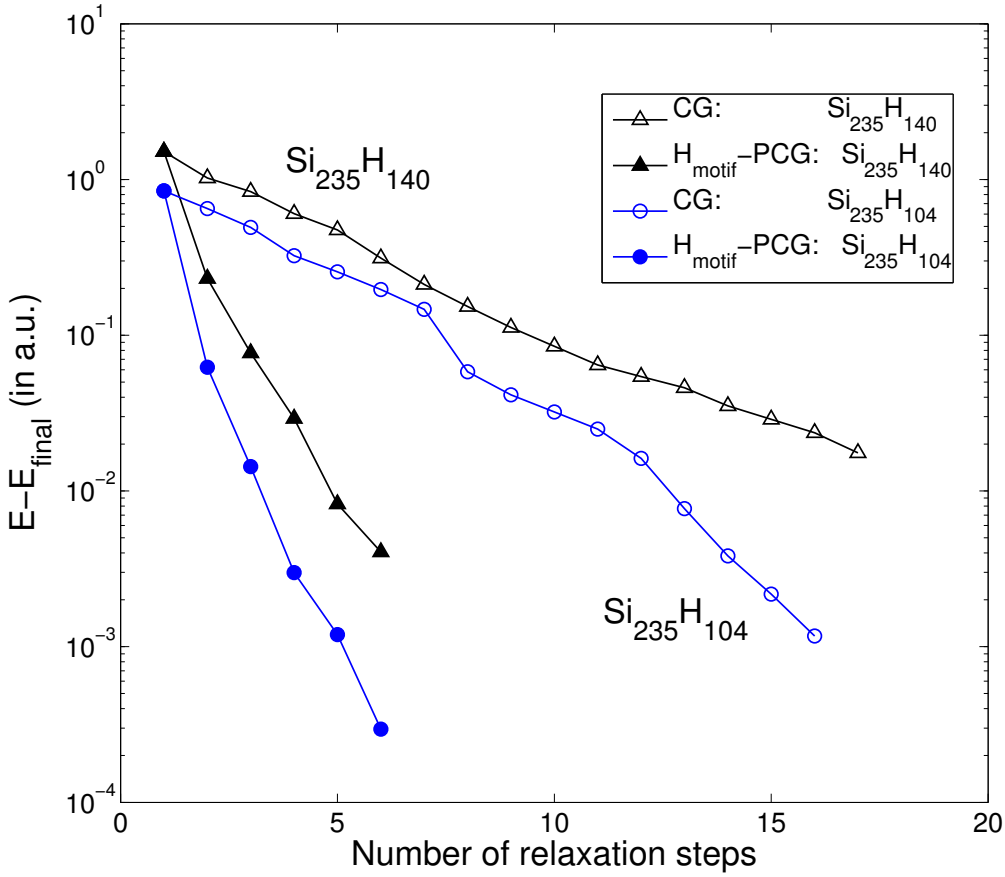

FIG. 4: The energy convergence in atomic relaxations for $Si_{235}H_{140}$ and $Si_{235}H_{104}$ quantum dots by applying CG (open diamond and open sphere, respectively), and the CG preconditioned by our motif Hessian H_{motif} (filled triangle and sphere, respectively) methods. In $Si_{235}H_{104}$ quantum dot, the surface Si atoms with two dangling bonds form Si-Si dimers (between two parallel zigzag chains in the (110) direction), this is a common reconfiguration on a Si surface.